

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representation of
The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Recognition and cleavage site of the intron-encoded *omega* transposase

(yeast/mitochondria/transposition/endonuclease)

L. COLLEAUX*, L. D'AURIOL†, F. GALIBERT†, AND B. DUJON*

*Département de Biologie Moléculaire, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France; and †Laboratoire d'Hématologie Expérimentale, Centre Hayem, Hôpital Saint Louis, F-75010 Paris, France

Communicated by Herschel L. Roman, March 31, 1988

ABSTRACT The optional group I intron of the mitochondrial 21S rRNA gene of *Saccharomyces cerevisiae* contains a 235-codon-long open reading frame the translation product of which (the *omega* transposase) catalyzes the formation of a double-strand break within the intron-minus (*omega*⁻) copies of the same gene. Purified *omega* transposase generates *in vitro* a 4-base-pair staggered cut with 3' hydroxyl overhangs at the exact position where the intron eventually inserts in the gene. Using randomly mutagenized synthetic oligonucleotides, single-base mutants were produced at 21 positions around the cleavage site. Experiments with these oligonucleotides show that the recognition site extends over an 18-base pair-long sequence within which minimal sequence degeneracy is tolerated. The intron-encoded *omega* transposase is, therefore, one of the most specific restriction endonucleases known to date.

The group I intron of the mitochondrial 21S rRNA gene of *Saccharomyces cerevisiae* (r1 intron), which is optional among different laboratory strains, shows the interesting property of propagating itself during crosses between intron-plus strains (*omega*⁺) and intron-minus ones (*omega*⁻) (1-4). The phenomenon, which is formally equivalent to a gene conversion, causes rapid spreading of this particular intron within yeast populations undergoing random mating. Insertion of the intron is associated with a coconversion of flanking exons extending over a few hundred base pairs (bp) on each side (5). Analysis of mtDNA in young zygotes of *omega*⁺ × *omega*⁻ crosses has revealed the transient formation of a double-strand break occurring at or near the recipient site of the intron-minus gene (*omega*⁻ site) (6, 7). Mutants that have an altered *omega*⁻ site do not show the double-strand break in similar crosses.

The r1 intron contains a 235-codon-long open reading frame (r1 ORF) that is conserved in a variety of yeast species belonging to the *Saccharomyces* and *Kluyveromyces* genera and shares distinctive features with ORFs of other group I introns (8, 9). Mutants within the r1 ORF easily demonstrate that its translation product must exist and is required for the gene conversion of the intron and its flanking exons (5, 10). Thus, the overall organization of the r1 intron is reminiscent of a transposable element encoding its own transposase.

To characterize the *omega* transposase, which, like the other intron-encoded proteins, remains undetectable in normal mitochondria, we have previously engineered the r1 ORF by modifying its nonuniversal codons (11). The resulting universal code equivalent directs, from expression plasmids, the synthesis of a protein with all characteristic properties expected of the genuine mitochondrial *omega* transposase. This protein exhibits a specific endonuclease activity, generating in *Escherichia coli* a double-strand break at *omega*⁻ sites carried on plasmids.

The *omega* transposase has already been partially purified from *E. coli* cells, and the fractions containing the partially purified substance can be shown to generate a double-strand cut within *omega*⁻ sites *in vitro* (12). Further purification to near homogeneity has now been achieved in our laboratory by following the *in vitro* activity of the protein.

We have now characterized the recognition and cleavage site of the intron-encoded *omega* transposase in detail using *in vitro* assays. The protein generates a 4-bp staggered cut at a specific sequence, leaving free 3' hydroxyl overhangs that can be religated. The recognition sequence extends over a continuous stretch of at least 18 bp almost exactly centered around the cleavage site.

MATERIALS AND METHODS

Strains and Plasmids. *E. coli* strain TG1 is $\Delta(lac-pro)$, *sup* E44, *thi* 1, *hsd* D5, *F'(traD36, lac* 19, *lac* Z Δ M15). *E. coli* growth media were made as described (13). Plasmid DNA preparations were made according to a reported procedure (14).

Construction of Randomly Mutagenized *omega*⁻ Sites. Approximately 60 pmol of complementary oligonucleotides (Fig. 1) were hybridized and ligated to reconstitute a 59-bp-long fragment containing the complete *omega*⁻ site (Fig. 2a). Oligonucleotides with free 5' termini internal to the duplex were phosphorylated before hybridization to allow ligation, whereas oligonucleotides with 5' termini external to the duplex were not phosphorylated to prevent multiple insertions. Each synthetic *omega*⁻ site was then ligated between the *Pst*I and *Xba*I sites of the M13 tg131 vector (Fig. 2b). The ligation mixture was then added with 200 μ l of TCM buffer (0.01 M Tris-HCl/0.01 M CaCl₂/0.01 M MgCl₂) and used to transfect CaCl₂-treated *E. coli* cells (strain TG1) diluted and plated on a lawn of TG1 cells to form plaques.

Analysis of Recombinant M13 Clones with Randomly Mutagenized *omega*⁻ Sites. Individual plaques were picked up at random and used to inoculate 2.5-ml cultures of TG1 cells in dYT medium to prepare both double- and single-stranded DNA as described by Messing (17). For each clone, single-stranded DNA was sequenced by the dideoxynucleotide chain-termination method (18), whereas the double-stranded DNA was tested for cleavage by a purified *omega* transposase (purification will be described elsewhere; A. Permin, unpublished work).

RESULTS

The *omega* Transposase Generates a 4-bp Staggered Cut with 3' Hydroxyl Overhangs at the Intron-Insertion Site. In yeast mitochondria as well as in transformed *E. coli* cells producing the *omega* transposase, the double-strand cut has been shown to occur at or near the exact intron-insertion point within the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ORF, open reading frame.

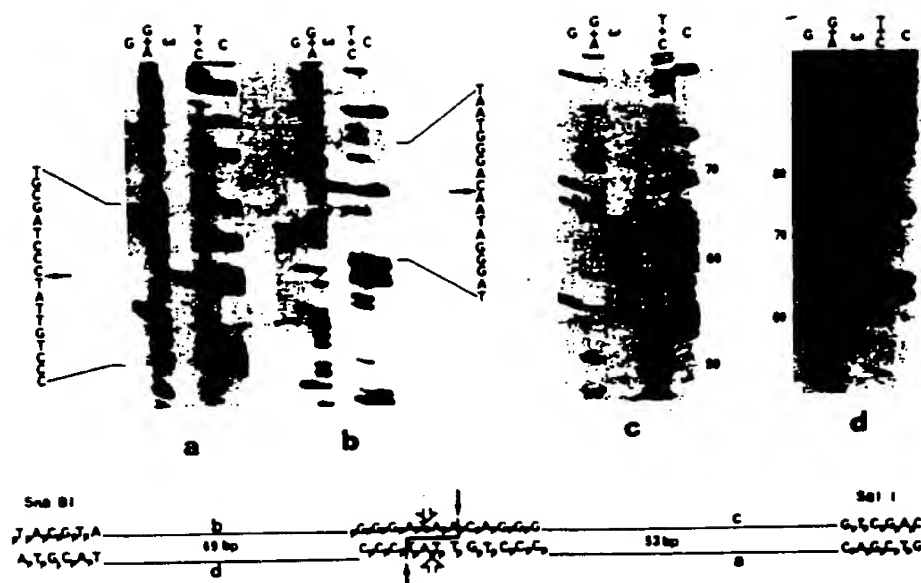


FIG. 3. Nature of the cleavage site. For mapping of the 3' termini: DNA from plasmid pSCM 522 (a pUC 12 derivative that contains the *omega*⁻ site) was restricted either at its unique *Sal* I site (a) or at its unique *Sna*BI site (b), dephosphorylated and 5' end-labeled with polynucleotide kinase and [γ -³²P]ATP. After secondary cleavage, the single end-labeled *Sal* I–*Sna*BI fragment was gel purified and then submitted to the action of the *omega* transposase. For mapping of the 5' termini: DNA from plasmid pSCM 522 was linearized at either its unique *Pst* I site (c) or at its unique *Sna*BI site (d) and 3' end-labeled with terminal deoxynucleotidyltransferase and [α -³²P]ddATP. Each end-labeled DNA was then submitted to the *omega* digestion. After incubation, the fragments were electrophoresed on an 8% polyacrylamide–urea sequencing gel calibrated using Maxam–Gilbert reactions of the same DNA fragments. The sequence is part of the 21S rRNA gene and shows the intron insertion point (↵) and the cleavage of each strand (†). The upper strand is identical to the 21S rRNA.

approximate the expected random distribution for an average of one substitution per oligonucleotide molecule. However, strong biases in distribution of the base substitutions were seen—some bases being mutated much more frequently than others. Part of this bias is probably from the transfection itself because results from independent transfection experiments from the same hybridizing mixture differed (data not shown). Finally, we isolated a total of 32 different single-base substitutions in the segment studied, some mutations being found several times among the 97 single mutants obtained. These selected mutants were plaque purified and sequenced again. Fig. 4 shows the sequence of all representative mutants. Mutations are available in 20 positions of a 25-bp segment. The *omega*^a, mutation previously isolated *in vivo* in yeast mitochondria (19) provides a 21st position.

Double-strand DNAs were purified from all single mutants as well as from some double mutants and were exposed to *omega* transposase action. Wild-type clones served as controls. Under our experimental conditions, digestion of the wild-type *omega*⁻ site is only partial (~50%) because the active fraction contains only a limited amount of *omega* transposase relative to the total amount of DNA used in the assays (A. Perrin, unpublished work). Cleavage efficiency at the mutant site of each clone was estimated by ethidium bromide fluorescence of fragments and referred to the efficiency at the wild-type site under similar conditions (Fig. 5). Three kinds of mutants can be recognized: those in which cleavage is undistinguishable from the wild type (+), those that are not cleaved at all (0), and those that are cleavable but with lower efficiency than the wild type (e). This last category contains a range of intermediate cleavages, some mutants (e.g., 357, 653, 711, 369, and 810) being only slightly affected, whereas others are severely affected. Comparisons of reactions performed with independent mutants having identical nucleotide sequences as well as independent reactions done

on the same mutant show the reproducibility of the experiment and the validity of this classification. We also tested the significance of the previous classification using three enzyme concentrations (1×, 3×, and 6×) resulting, respectively, in cleavage of 50%, 80%, and 99% of the wild-type site. Under such conditions, a mutant belonging to the 0 class (318) remained completely uncleaved, whereas two mutants belonging to the e class (357 and 840) showed significantly reduced cleavage (5%, 25%, and 50% for, respectively, the 1×, 3×, and 6× enzyme concentrations).

The three kinds of mutants are indicated in Fig. 4, which shows the following: (i) Mutants affecting cleavage occur from position -7 (mutant 155) to position +11 (mutant 147), indicating that the recognition site of the *omega* transposase extends at least over 18 bp. The absence of mutants at position -8 and after position +11 does not permit any more precise limits on the recognition site; (ii) recognition of the *omega*⁻ site by *omega* transposase involves a continuous stretch within this 18-bp segment because at least one mutant exists at every position (except +9 and +10) of the 0 or e type; (iii) within this 18-bp segment very few mutations (826 and 516) do not affect cleavage.

DISCUSSION

Our results demonstrate that an intron-encoded protein catalyzes the formation of a site-specific double-strand break within a gene, leaving 4-bp sticky ends with 3' hydroxyl overhangs like those produced by class II restriction endonucleases (20). The position of the break coincides exactly with the intron-insertion site, indicating it as the initiating event of the whole intron-insertion process. Thus the organization of this intron is reminiscent of a transposable element encoding its own transposase (21)—except that the

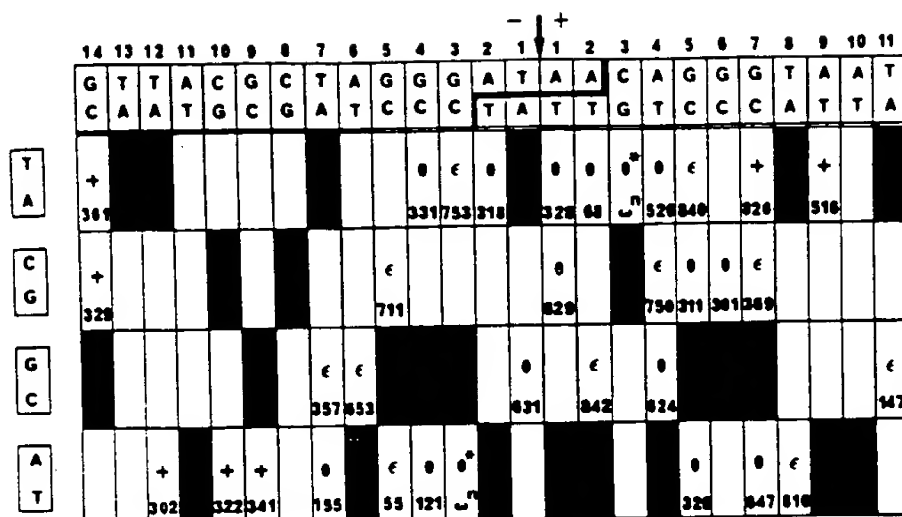


FIG. 4. Mutational analysis of the recognition site. Sequence of the 21S rRNA gene around the ω site is given (Top). The arrow indicates the intron insertion point in the middle of the staggered cut (thick line). The three possible bp substitutions (left boxes) at every position are shown; the black boxes correspond to wild-type sequence. Numbers refer to mutants and ω refers to mutants isolated *in vivo* in yeast (19). Efficiency of cleavage is as defined in Fig. 5: +, mutant cleaved as wild-type by the ω transposase; 0, no cleavage by the ω transposase; ϵ , reduced level of cleavage by the ω transposase. For the two ω mutants, the cleavage efficiency (0) has been deduced from *in vivo* experiments in yeast (0%). Mutants 121 and 520 have the following alterations in addition to the one indicated: 121, a 1-bp deletion at position -14; 520, a A \rightarrow T transversion at position -15.

latter intron always inserts in the same recipient site and does not duplicate bases at that site.

The described phenomenon is analogous in several ways with the yeast mating type interconversion catalyzed by the HO endonuclease (22) and can be interpreted by the double-strand break-repair model (23, 24). However, for the ω system, degradation of preexisting sequences around the cleavage site is not required. It follows that one free 3' extension generated by ω transposase can be a primer to elongate a copy of the intron sequence as shown in Fig. 6. This model differs from the classic double-strand break-

repair model by the facts that (i) this model requires special ends, (ii) recombinational intermediates include the ω transposase, and (iii) a single-strand intermediate exists in this model.

Our results also demonstrate that the recognition site of the ω transposase extends over at least 18 bp. Most base substitutions within these 18 bp either completely abolish cleavage or significantly reduce cleavage efficiency, indicating that the ω transposase tolerates minimal sequence degeneracy. The 29 single-base mutants isolated from the 54 possible ones (54%) span 17 positions of the 18 bp forming the recognition site. Because no selection was applied in the protocol, it is reasonable to assume that type distribution of the remaining 25 possible mutants would resemble that already found. Thus, the ω endonuclease is much more specific than any other bacterial restriction endonuclease known so far (26). The recognition site of the HO endonuclease also extends over an 18-bp-long segment, but the degree of degeneracy tolerated is less precisely known due to fewer available mutants (27). Interestingly, the only other eukaryotic double-strand endonucleases known to date, all from yeast (HO, Sce I, and Sce II), generate 4-bp extensions with 3' overhangs (22, 28). Yet, the sequences cleaved by these proteins do not obviously resemble one another or the ω site. Thus, the ω transposase may eventually become a useful restriction enzyme for specific experiments such as genomic mapping and cloning.

Mutations within the recognition site that reduce or abolish cleavage can be interpreted in either of two ways: alteration of the binding of the protein to the site or binding without cleavage. Preliminary experiments using one of the ω mutants favor the second possibility (A. Perrin, unpublished work). The large size of the recognition site does not mean that the ω transposase, which is a relatively small protein (235 amino acids), actually contacts all 18 bp of the recognition site. Mutations at some positions may alter the conformation of the helix at nearby positions involved in the contacts. Precise protein-DNA interactions need to be studied. The λ repressor (29), the λ integrase (30), or the resolvase

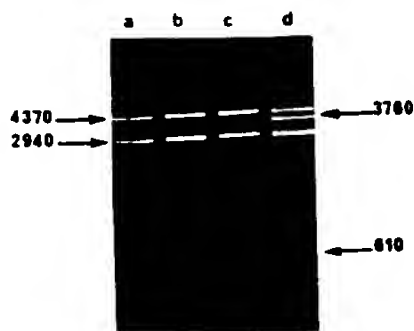


FIG. 5. Assays for cleavage by the ω transposase. Assays were done in standard restriction buffer: 0.02 M Tris-HCl/6 mM MgCl₂/0.05 M NaCl/1 mM DL-dithiothreitol, pH 7.5, containing bovine serum albumin at 100 μ g/ml. The double-stranded DNA was first digested by Cla I for 1 hr and then treated with 2-5 μ l of purified fractions of ω transposase for another hour at 37°C. Sizes of the Cla I restriction fragments of the M13 recombinants are indicated in bp on the left of the gel; fragment sizes generated after cleavage by the ω transposase are indicated on the right. The latter correspond to the cleavage, at the ω site, of the 4370-bp-long fragment. Cleavage efficiency was estimated from the ratio of the 3760-bp fragment to the 4370-bp fragment. Lanes: a and c, mutants of the 0 type; b, mutant of the ϵ type; and d, mutant of the + type.

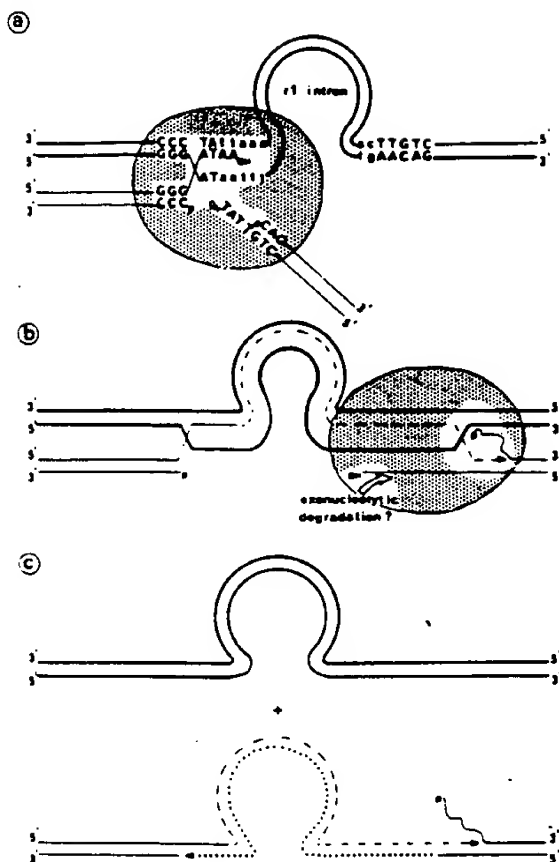


FIG. 6. Hypothetical mechanism of intron insertion based on the nature of the cut. The figure stresses that the 4-bp extension (5'ATAA_{OH}) can serve directly as primer (a) for elongation of the top strand by DNA polymerase after a strand-assimilation process similar to that mediated by RecA protein in *E. coli* (25) (b; broken line). The top strand is identical to the 21S rRNA. When the newly synthesized strand reaches the downstream exon it can, in turn, be assimilated by the corresponding sequence from the split ω -duplex (b) and serves as matrix for the elongation of the bottom strand (c). This elongation, however, requires the exonucleolytic degradation of, at least, the 4-bp extension (5'TTAT_{OH}) on the bottom strand. Resolution of the intermediate could be either by branch migration (as shown), leaving an intact ω -duplex, or by cleavage of the two Holliday junctions, forming a hybrid between the old and the new strand. Lowercase letters and uppercase letters represent sequences of the r1 intron and of the exons, respectively. Shadowed section represents a hypothetical complex that holds the split ω -duplex together with the ω -duplex; the ω transposase could be contained in such a complex.

family (31) probably offer better analogies for the ω transposase-DNA interactions than bacterial type II restriction endonucleases. The ω transposase, however, differs from these models for at least three reasons: (i) the ω site is nonsymmetrical; (ii) secondary binding sites are unlikely (*in vitro* cleavage occurs even if the ω site is transported into a completely new environment); and (iii) no recombinogenic activity of the ω transposase could be found in experiments involving replicons with two ω sites (12).

The extreme sequence specificity shown by the intron-encoded ω transposase is puzzling with respect to the

small size of the mitochondrial genome and what advantage, if any, the yeast gains by synthesizing such a specific endonuclease. If the role of the intronic protein were to propagate the intron by introducing it at new genetic locations, then too high a specificity could only limit efficiency of the process. On the other hand, synthesizing double-strand endonucleases with a relatively low specificity is obviously deleterious for a cell without the corresponding modification mechanism. Bacteria have evolved sophisticated restriction-modification systems, whereas eukaryotes have not. Perhaps we should not be surprised that the few eukaryotic endonucleases characterized to date are so specific.

We thank C. Monteilhet and A. Perrin for the generous gift of purified fractions of the ω transposase. We thank them as well as Alain Jacquier, François Michel, Sophie Moutierac, and Christine Perez for invaluable discussions and comments. L.C. is the recipient of a student fellowship from Ecole Normale Supérieure. This work was performed in part while L.C. and B.D. were at the Centre de Génétique Moléculaire du Centre National de la Recherche Scientifique (CNRS). This work was supported by Grants from Institut National de la Santé et de la Recherche Médicale 861007 and from University Pierre and Marie Curie to B.D. and by a grant from CNRS Laboratoire d'Oncologie et d'Immunohématologie to F.G.

1. Michel, F., Jacquier, A., & Dujon, B. (1982) *Biochimie* 64, 867-881.
2. Michel, F., & Dujon, B. (1983) *EMBO J.* 2, 33-37.
3. Bolotin, M., Coen, D., Deutsch, J., Dujon, B., Netter, P., Petrochilo, E., & Slonimski, P. P. (1971) *Bull. Inst. Pasteur* 69, 215-239.
4. Dujon, B., Bolotin-Fukuhara, M., Coen, D., Deutsch, J., Netter, P., Slonimski, P. P., & Weil, L. (1976) *Mol. Gen. Genet.* 143, 131-165.
5. Jacquier, A., & Dujon, B. (1985) *Cell* 41, 383-394.
6. Dujon, B., Cortarel, G., Colleaux, L., Betermier, M., Jacquier, A., d'Auriol, L., & Galibert, F. (1985) *Perspectives and Achievements in Mitochondrial Research*, ed. Palmieri, F. (Elsevier, Amsterdam), pp. 215-225.
7. Zinn, A. R., & Butow, R. A. (1985) *Cell* 40, 887-895.
8. Dujon, B. (1980) *Cell* 20, 185-197.
9. Jacquier, A., & Dujon, B. (1983) *Mol. Gen. Genet.* 192, 487-499.
10. Macreadie, I. G., Scott, R. M., Zinn, A. R., & Butow, R. A. (1985) *Cell* 41, 395-402.
11. Colleaux, L., d'Auriol, L., Betermier, M., Cortarel, G., Jacquier, A., Galibert, F., & Dujon, B. (1986) *Cell* 44, 521-533.
12. Dujon, B., Colleaux, L., Jacquier, A., Michel, F., & Monteilhet, C. (1986) in *Extrachromosomal Elements in Lower Eukaryotes*, eds. Wickner, R. B., Hinnebusch, A., Lambowitz, A. M., Gunsakus, I. C., & Hollaender, A. (Plenum, New York), pp. 5-27.
13. Miller, J. H. (1972) *Experiments in Molecular Biology* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
14. Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
15. Gait, M. J. (1984) *Oligonucleotide Synthesis: A Practical Approach*, eds. Rickwood, D., & Hames, B. D. (IRL, Oxford, UK).
16. Maxam, A., & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
17. Messing, J. (1983) *Methods Enzymol.* 101, 20-78.
18. Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
19. Dujon, B., & Jacquier, A. (1983) in *Mitochondria*, eds. Schreyer, R. J., Wolf, K., & Kaudewitz, F. (de Gruyter, Berlin), pp. 389-403.
20. Modrich, P., & Roberts, R. J. (1982) in *Nucleases*, eds. Lim, S. M., & Roberts, R. J. (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 109-154.
21. Shapiro, J. (1983) *Mobile Genetic Elements* (Academic, New York).
22. Kostriken, R., Strathern, J. N., Klar, A. J. S., Hicks, J. B., & Heffron, F. (1983) *Cell* 35, 167-174.
23. Orr-Weaver, T. L., Szostak, J. W., & Rothstein, R. J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6354-6358.
24. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., & Stahl, F. W. (1983) *Cell* 33, 25-35.
25. Flory, S. S., Tsang, J., Muniyappa, K., Bianchi, M., Gonda, G., Kahn, A., Azhdar, E., Egner, C., Shaner, S., & Radding, C. M. (1984) *Cold Spring Harbor Symp. Quant. Biol.* 49, 513-523.
26. Roberts, R. J. (1987) *Nucleic Acids Res.* 15, Suppl. r189-r217.
27. Nickoloff, J. A., Chen, E. Y., & Heffron, F. (1986) *Proc. Natl. Acad. Sci. USA* 83, 7831-7835.
28. Watabe, H., Iino, T., Kaneko, T., Shibata, T., & Ando, T. (1983) *J. Biol. Chem.* 258, 4663-4665.
29. Plashne, M. (1986) *A Genetic Switch: Gene Control and Phase A* (Blackwell, Oxford, UK).
30. Craig, N. L., & Nash, H. A. (1983) in *Mechanism of DNA Replication and Recombination*, ed. Cozzarelli, N. R. (Liss, New York), Vol. 10, pp. 617-636.
31. Krasnow, M. A., & Cozzarelli, N. R. (1983) *Cell* 32, 1313-1324.